Experimental and bioinformatic characterisation of the promoter region of the Marfan syndrome gene, FBN1

Genomics 94 (2009) 233–240

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Experimental and bioinformatic characterisation of the promoter region of the Marfan syndrome gene, *FBN1*

Kim M. Summers [a,b,*], Nilesh J. Bokil [c], John M. Baisden [b], Malcolm J. West [d], Matthew J. Sweet [c], Liza J. Raggatt [c,e], David A. Hume [a,c]

[a] The University of Edinburgh, The Roslin Institute and Royal (Dick) School of Veterinary Studies, UK
[b] The University of Queensland, School of Chemistry and Molecular Biosciences, Australia
[c] The University of Queensland, Institute for Molecular Bioscience, Australia
[d] The University of Queensland, School of Medicine, Australia
[e] The University of Queensland, Centre for Clinical Research, Australia

ARTICLE INFO

Article history:
Received 6 March 2009
Accepted 19 June 2009
Available online 30 June 2009

Keywords:
Marfan syndrome
Fibrillin-1
Gene expression
Dominant genetic conditions
Reporter genes
Promoter regions
Mesenchyme

ABSTRACT

Mutations in the *FBN1* gene, encoding the extracellular matrix protein fibrillin-1, result in the dominant connective tissue disease Marfan syndrome. Marfan syndrome has a variable phenotype, even within families carrying the same *FBN1* mutation. Differences in gene expression resulting from sequence differences in the promoter region of the *FBN1* gene are likely to be involved in causing this phenotypic variability. In this report, we present an analysis of *FBN1* transcription start site (TSS) use in mouse and human tissues. We found that transcription of *FBN1* initiated primarily from a single CpG-rich promoter which was highly conserved in mammals. It contained potential binding sites for a number of factors implicated in mesenchyme differentiation and gene expression. The human osteosarcoma line MG63 had high levels of *FBN1* mRNA and secreted fibrillin-1 protein to form extracellular matrix fibres. The human embryonic kidney line HEK293 and two breast cancer lines MCF7 and MDA-MB-231 had levels of *FBN1* mRNA 1000 fold lower and produced negligible amounts of fibrillin-1 protein. Therefore MG63 appears to be the optimal cell line for examining tissue-specific, biologically relevant promoter activity for FBN1. In reporter assays, the conserved promoter region was more active in MG63 cells than in non-*FBN1*-expressing lines but additional elements outside the proximal promoter are probably required for optimal tissue-specific expression. Understanding the regulation of the *FBN1* gene may lead to alternative therapeutic strategies for Marfan syndrome.

Introduction

Freely available large scale genome-wide information can be used to investigate the evolution, expression and variation of human disease genes. Marfan syndrome (OMIM 154700, OMIA 1204) is a connective tissue disease, inherited in an autosomal dominant fashion, with high penetrance but considerable phenotypic variability (for example, see [1,2]). It is generally caused by mutation of the *FBN1* gene, encoding the microfibrillar protein fibrillin-1. Severely affected individuals have overgrowth of the long bones, cardiovascular abnormalities including aortic dilatation, ocular problems including subluxation of the lenses and abnormalities of eye shape, reduced subcutaneous fat and muscle tissue and abnormalities of the lungs, skin and integument [3]. Thus Marfan syndrome affects tissues of mesenchymal origin, such as muscle, adipose and bone. Mutation carriers within a family can present with highly variable clinical manifestations, and genetic counselling based on discovery of a mutation is difficult because of this variability [4]. Recent studies [5–7] are consistent with a haploinsufficiency model for dominance of *FBN1* mutations, suggesting that in at least some cases reduction in the level of normal fibrillin-1 protein by 50% is sufficient to cause the phenotype. Alterations 5′ of the coding sequence, which could affect gene transcription levels, have been identified in Marfan patients [8]. Levels of gene expression are heritable and can contribute to disease [9,10] and *FBN1* promoter region variability could thus contribute to phenotypic variability within and between families [5].

Early analysis of expressed sequences in human mRNA showed that *FBN1* transcripts begin with untranslated sequence derived from one of three alternative 5′ exons (named in order Exon B, A and C), spliced to the first exon with verified coding sequence (originally named Exon M, now Exon 1) [11]. A recent study [12] showed that conserved sequences upstream of human Exon A conferred promoter activity in transient transfection analysis using HEK293 and HT1080

Abbreviations: bp, base pairs; CAGE, cap analysis of gene expression; DPE, downstream promoter element; DMEM, Dulbecco's modified Eagle's medium; FANTOM, Functional Annotation of the Mammalian Genome; HEK, human embryonic kidney; Inr, initiator element; MEM, minimal essential medium; OMIA, Online Mendelian Inheritance in Animals; OMIM, Online Mendelian Inheritance in Man; RACE, rapid amplification of cDNA ends; TSS, transcription start site.

* Corresponding author. The Roslin Institute, Roslin, Midlothian EH25 9PS, UK. Fax: +44 131 440 0434.
E-mail address: kim.summers@roslin.ed.ac.uk (K.M. Summers).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.06.005

cells. Other authors [13] have speculated that the initiation of transcription may occur upstream of the most 5′ putative exon, Exon B. Since differences in *FBN1* expression may contribute to phenotypic variability in Marfan syndrome [5,6], it is important to understand the initiation and regulation of transcription of this gene. We have used extensive bioinformatic resources to survey the human *FBN1* gene and identify the major transcription start site. We analysed *FBN1* expression in a human osteosarcoma cell line, chosen because of the high level of *FBN1* expression in bone and primary osteoblast cultures of human [14] and mouse (BioGPS *FBN1* entry) and the bone phenotype in Marfan syndrome [3]. Our studies illustrate the power of available genomic resources in the study of transcriptional regulation, and provide a framework for investigating the molecular basis of phenotypic variation in Marfan syndrome patients.

Results

Structure of the FBN1 promoter region

The FANTOM (Functional Annotation of the Mammalian Genome) project is an international collaborative effort to map transcriptional start sites (TSS) of mouse and human, sponsored by the RIKEN Omics Institute in Japan. Using CAGE (cap analysis of gene expression), a form of genome-wide 5′-RACE analysis which detects the initial bases of a transcript, FANTOM3 identified over 11 million transcripts in mouse (over 7 million mapped to unique sites in the genome) and nearly 6 million in human (more than 3 million mapped) [15]. We used this resource to find the TSSs of the human and mouse *FBN1* genes. For mouse, a total of 788 CAGE tags (comprising the first 20 bases of transcripts) mapped across the *Fbn1* gene (Supplementary Fig. 1A). For human, 966 tags were detected across the region of *FBN1* (Supplementary Fig. 1B). There were clusters of CAGE tags indicating TSSs in the forward orientation over nearly 1 kb encompassing Exons A, B and C in both human (719 tags) and mouse (612 tags) (Fig. 1). Many of these tags were singletons that did not overlap with each other to form a cluster.

Against a background of low-level initiation across the region (as shown in Fig. 1 and Supplementary Fig. 1A), there were two start site clusters 5′ of Exon 1 in mouse (Fig. 1). One, with 484 CAGE tags, corresponded with Exon A. The other start site cluster, with 85 tags, was located closer to Exon 1 and corresponded with the published Exon C of human (Fig. 1). In addition 4 tags were close to the position of Exon B of human (Fig. 1). These results indicate that the promoter initiating at Exon A was most commonly used in the tissue types examined in mouse. The CAGE results also provided information on cell types which express *FBN1* mRNA. In mouse, the majority of detected transcripts (440 of 484 for Exon A and 45 of 85 for Exon C) came from embryo mRNA. Mouse RNA also showed two additional start sites (with 18 and 12 tags respectively) in the reverse direction in the second intron (shown in Supplementary Fig. 1A). The sites were in highly conserved sequences.

For human, there was a single major TSS cluster, corresponding with Exon A. 373 of the 658 transcripts were found in mRNA prepared from skin cells and 177 were from adipose tissue. There was no focus of TSS in the region of either Exon B or Exon C, with additional TSS clusters of one to 13 tags scattered over the whole region from Exon B to Exon 1 (Fig. 1). This indicates that, for the tissues and cell types sampled, there was no selective use of Exons B and C in humans. The reverse TSSs of the mouse second intron were not seen in the human RNA but there was a cluster of 19 forward tags at the beginning of coding Exon 8 (Supplementary Fig. 1B). The differences between human and mouse may be attributable to the different tissues

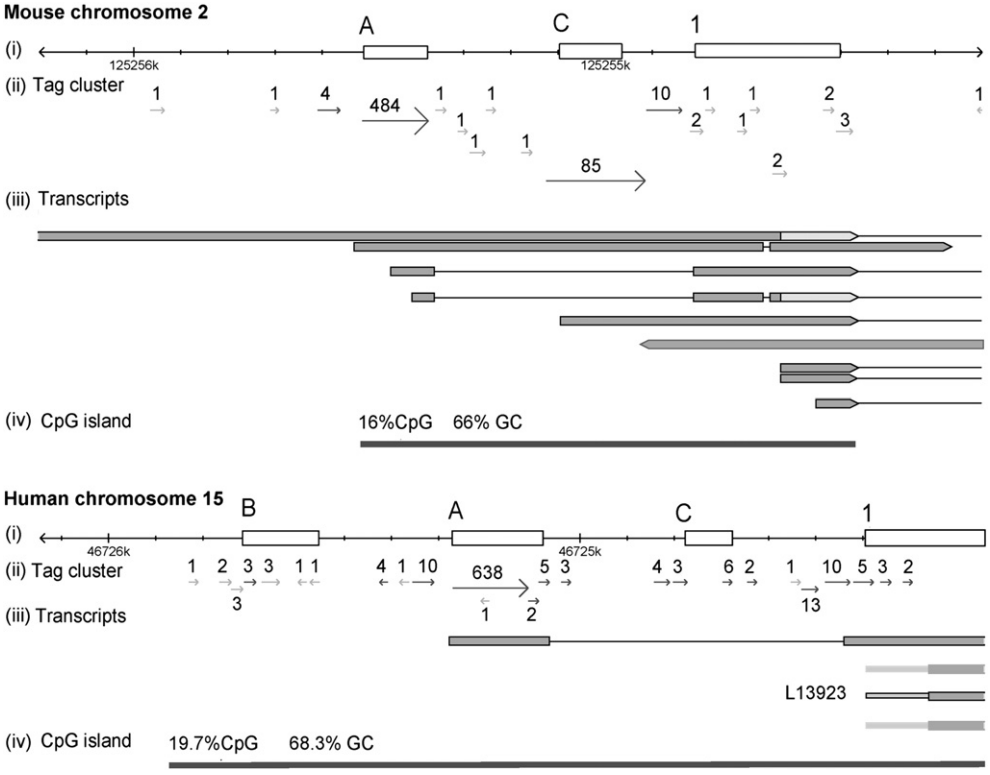

Fig. 1. Transcription start sites of mouse and human *FBN1* genes, detected by CAGE analysis. The diagrams are redrawn from those generated by the FANTOM3 Genome Elements Viewer as at March 2009 and show distribution of transcription start sites for mouse (upper) and human (lower) *FBN1* genes 5′ of the putative start codon in Exon 1. Line (i) shows the base pair position on the physical map of the chromosome for mouse (mm5) and human (hg17). The positions of putative Exons A, B and C and the first coding exon, Exon 1, are shown by boxes. In line (ii) arrows show the location and placement of TSS clusters. Numbers indicate the number of TSSs found at that location and the length of the arrow shows the region covered by overlapping tags in the cluster. The direction of the arrow shows the direction of transcription. Gray arrows show low frequency tags (less than 1 in 10^5 total tags). Line (iii) shows validated transcripts. Boxes indicate exons and lines indicate sequences which are spliced out. Line (iv) shows the regions of CpG islands. TSSs covering the complete mouse and human *FBN1* genes are shown in Supplementary Fig. 1.

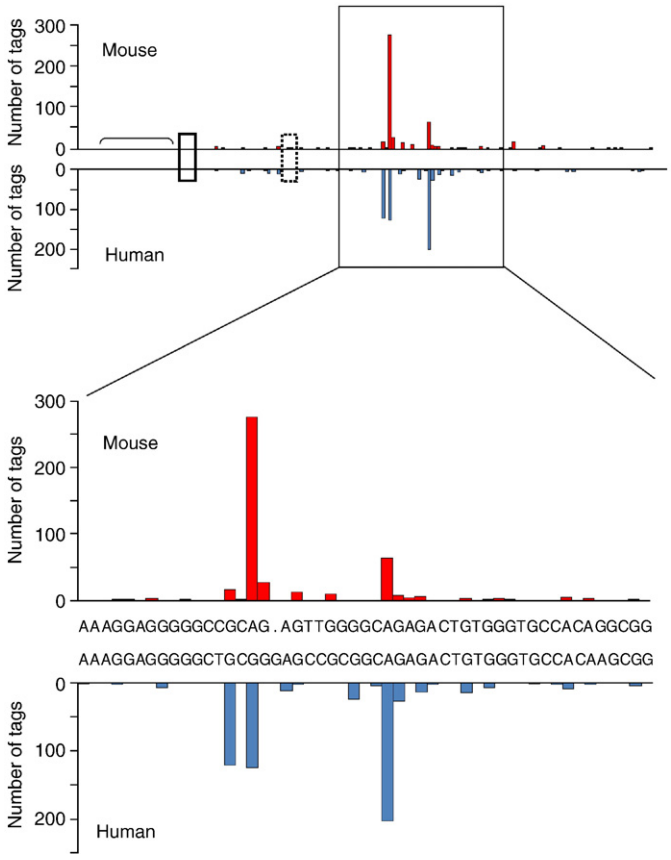

Fig. 2. Histograms of mouse and human TSS usage. Mouse and human sequences were aligned and the number of tags starting on each base was graphed. *Y* axis indicates number of transcripts starting on that base. Bracket shows the pyrimidine region. Small boxes shows the putative Inr (solid) and DPE (dotted) of Guo et al. [12]. Large box indicates the region expanded below, showing the first base of the transcripts. Main sequence covers bases −1046 to −856 in the human sequence, numbered from the ATG start codon (at position 5348 in GenBank entry L19896).

sampled in the CAGE analysis, or may reflect a species-specific difference in regulation of *FBN1*.

We analysed the nucleotides at which the transcripts initiated for human and mouse (Fig. 2). The large majority conformed to a Py Pu initiator consensus, with CA the most common dinucleotide (A being the first base of the transcript). Although Exon A transcripts initiated over a wide region, in human the majority initiated at one of three sites, while in mouse there was a dominant TSS with a second also showing increased starts. The alignment of the CAGE tags on the sequence of human and mouse (Fig. 2) revealed that base changes between species were correlated with differences in the frequency of starting at particular bases. A GC-rich element likely to bind the transcription factor SP1 started 40 bases upstream of the first major TSS in both species. A series of GC-rich elements conforming to the variant GCG trinucleotide repeat downstream promoter element identified previously by genome-wide promoter analysis [16] was found 3′ of the major TSSs (shown in Supplementary Fig. 2). There was an apparent deletion in rodents bringing the GCG echo closer to the start site (Supplementary Fig. 2). The region of the TSSs of *FBN1* conformed to the definition of a CpG island (lines labelled iv in Fig. 1) [17].

Analysis of conserved FBN1 sequence

Sequence conservation of the region around the Exon A TSS was examined in 28 mammalian species for which adequate sequence information was available in the Ensembl data base and trace server (species listed in Supplementary Fig. 2). Transcription factor binding motifs were detected using rVista (TRANSFAC motif library) and compared with analysis using the JASPAR motif library (see Materials and methods). Table 1 shows conserved sites in the region of the TSSs and Supplementary Fig. 2 shows the same sites across all species tested. A highly conserved element (TTTCTCGCGAGAAA), containing the palindrome reported previously [12], was identified to contain an inverted repeat dual binding motif for the transcription factor E2F [18,19]. Across 28 mammals, the putative E2F binding sites only varied by deletion of the terminal T and A in *Echinops telfairi* and addition of a CA dinucleotide in *Myotis lucifugus* (Supplementary Fig. 2). Because of the repeat nature of the sequences, and the fact that these sequences were derived from the trace database and are not fully validated, these could be sequencing errors, but may represent true variation. Inverted repeats of a core sequence AAGTAT (Table 1 and Supplementary Fig. 2) were also conserved. These sites were candidate GATA motifs. Two conserved E-boxes (CAGCTG) were detected (Table 1 and Supplementary Fig. 2) with a third site further downstream in mouse and rat only. Three conserved E-boxes were also found in the Exon C promoter region (not shown). As indicated in Supplementary Fig. 2, 3′ of the main initiator region there were three conserved binding motifs for members of the RUNX transcription factor family (TGTGG) [20].

Immediately upstream of the major TSS region was a stretch of pyrimidines (length ranging from 21 in *Oryctolagus cuniculus* to 39 in *Spermophilus tridecemlineatus*) flanked by a string of As at the 3′ end and a G-rich region at the 5′ end (Supplementary Fig. 2). The 5′ end of the G-rich region contained a conserved octamer-like motif (TGTCATTT) which could potentially bind POU-domain factors

Table 1
Transcription factor binding motifs identified in the conserved *FBN1* promoter region.

Motif sequence	Location[a]	Transcription factor families	Tissue/function	Reference
GTGGAAAG	−1107 to −1114	C/EBP	Cell cycle, differentiation in multiple lineages, activation of RUNX2 (CEBPD)	[35,36]
TGTCATTT	−1051 to −1044	POU-domain factors	General transcription factor, mesenchymal stem cell lineage determination (OCT4)	[37,38]
TTTTCTTTTTCTTTCTTTTTTT	−1046 to −1025	Broad complex, paired homeobox, forkhead and HMG-IY families	Mesenchyme and other lineages, myogenesis (PAX3/7)	[39]
TTTTTTTAAAAAAA	−1031 to −1018	MEF2, homeobox factors e.g. PRX family	Muscle development Mesenchyme differentiation	[40,41]
AAGTAT	−1010 to −1005	GATA	Mesenchymal lineage determination (GATA4)	[42,43]
ATACTT	−969 to −964	GATA	Mesenchymal lineage determination (GATA4)	[42,43]
TTTCTC	−1004 to −999	E2F	Proliferating cells; inhibition of chondrocyte differentiation (E2F1)	[18,19,44,45]
GAGAAA	−994 to −989	E2F	Proliferating cells; inhibition of chondrocyte differentiation (E2F1)	[18,19,44,45]
TGTGG	−918 to −914	RUNX	Skeletal development (RUNX2)	[20,46]
CCACA	−909 to −905	RUNX	Skeletal development (RUNX2)	[20,46]
CCACA	−893 to −889	RUNX	Skeletal development (RUNX2)	[20,46]
CAGCTG	−890 to −885	MYO-D, MYOGENIN, SNAIL	Muscle development	[47,48]
CAGCTG	−881 to −876	MYO-D, MYOGENIN, SNAIL	Muscle development	[47,48]

[a] Location in the human sequence; bases 5′ of the ATG start codon at position 5348 in GenBank entry L19896.

Table 2
FBN1 mRNA expression levels in cell lines.

Cell line	Expression relative to HPRT (SD)		
	Experiment 1	Experiment 2	Experiment 3
MG63	14.07 (4.67)	65.34 (5.01)	21.86 (5.36)
MDA-MB-231	0.21 (0.03)	0.20 (0.03)	NA
MCF7	NA	0.20 (0.03)	0.21 (0.01)
HEK293	0.01 (0.0006)	0.06 (0.01)	0.01 (0.001)

Results for representative experiments are shown. For each, cDNA synthesis and quantitative PCR were carried out at the same time for duplicate RNA samples, each assayed in triplicate. NA — not assayed in the experiment.

(Table 1). This sequence was invariant in the 28 mammalian species (Supplementary Fig. 2). The 3′ string of Ts followed by a string of As was predicted to be a binding site for the myogenic enhancer-2 (MEF-2) family. The AT-rich sequence also contained predicted binding motifs for a range of homeobox factors. The conserved pyrimidine string flanked by purines was not found in the 5′ regions of other members of the fibrillin/LTBP family or other genes. There were a number of potential transcription factor binding motifs within this sequence, including motifs for the broad complex, paired homeobox, forkhead and HMG-IY families. In addition, analysis using the RECON programme [21] suggested that the pyrimidine-rich sequence was likely to have a role in nucleosome phasing (data not shown).

Despite the remarkable conservation of the promoter region in mammals (also noted in other reports [11–13] using a more limited range of species), there were no matches to the Exon A sequence or the TSS region found by BLAT or BLAST analysis in completed or draft genomes of any other vertebrate species, including birds (chicken and zebra finch), fish (fugu, zebrafish, medaka, stickleback, pufferfish), a reptile (anole lizard) and an amphibian (*Xenopus tropicalis*), nor in the platypus or the wallaby *Macropus eugeneii*. Many of these species do have annotated *FBN1* genes, although the initial exons are frequently missing. For example, the first exon annotated for chicken corresponds with the second coding exon of human and the first for platypus with the sixth coding exon. The absence of this sequence in some mammals and in other vertebrates may thus be due to incomplete sequencing, or may reflect the evolutionary relationship of the Australian fauna to non-mammalian vertebrates.

As can be seen in Supplementary Fig. 2, sequence homology across species degenerates 3′ of the strong Exon A splice donor (CCA/GT) that is conserved across species and around 300 bp upstream of the pyrimidine stretch, before Exon B. BLAT and BLAST analysis of the previously identified human Exon B (~300–400 bp upstream of Exon A) and Exon C (~300–400 bp downstream of Exon A) found no significant matches in the mouse genome.

FBN1 mRNA and protein expression in cultured cells

Analysis of public microarray data (see Materials and methods) showed that *FBN1* was strongly expressed in human adipocytes, smooth muscle, placenta, cardiac myocytes and some fetal tissues. From the microarray results, expression was low in most cell lines, other than those of mesenchymal origin. In the mouse, *Fbn1*

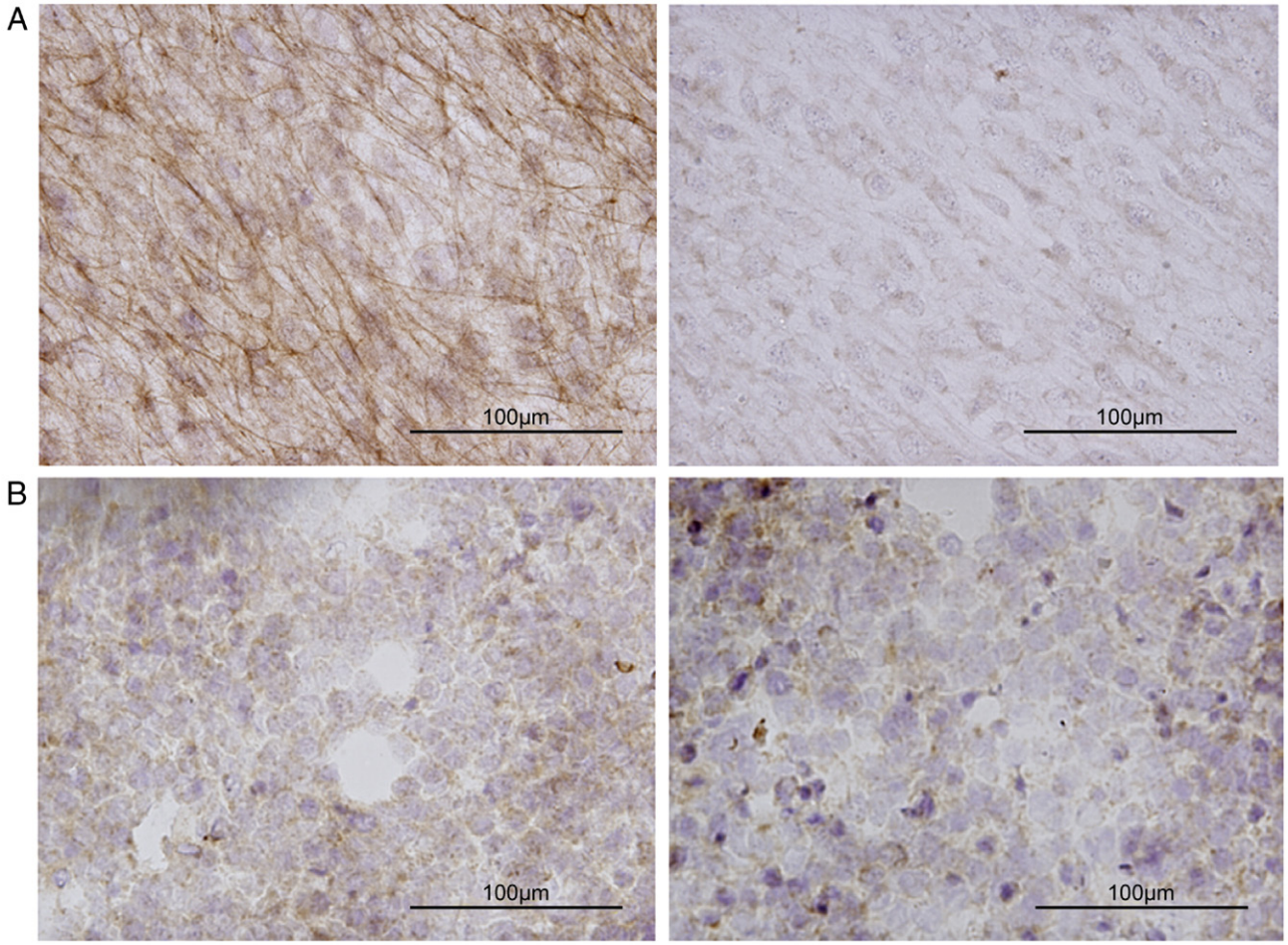

Fig. 3. Immunocytochemical staining of cell lines for fibrillin-1. All cells were stained at day 14 after plating. Left hand panels show staining using a mouse monoclonal anti-fibrillin-1 antibody; right hand panels show staining using an isotype antibody control (see Materials and methods). Images are representative of at least three separate experiments. (A) MG63; (B) HEK293. Original magnification 60×.

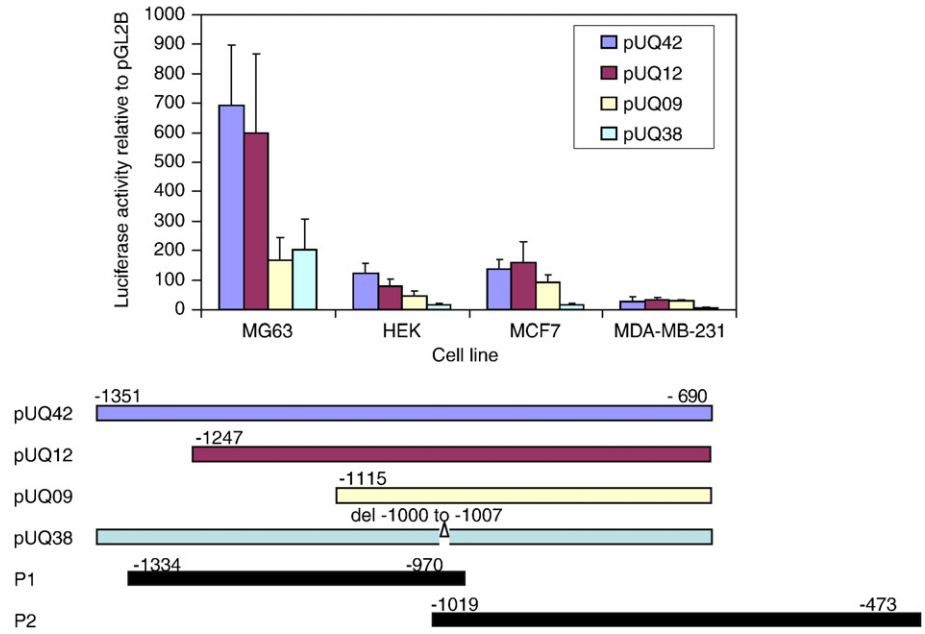

Fig. 4. Luciferase activity for different cell lines and promoter constructs. *Y* axis shows luciferase activity relative to the empty vector pGL2B. Results show average of five (HEK293), four (MG63, MCF7), or two (MDA-MB-231) separate experiments. Bars show standard error. Values for pGL2C (SV40 promoter control) were 1990 ± 1219 (MG63); 25 ± 10 (HEK293); 1238 ± 370 (MCF7) and 446 ± 156 (MDA-MB-231). Horizontal lines below show the relationship of the *FBN1* sequence contained in each construct, compared with the P1 and P2 constructs used previously [12].

expression was also high in tissues and cells of mesenchymal origin. To confirm the level of expression of human *FBN1* in the cell lines used in the present study, we surveyed mRNA levels using quantitative reverse transcriptase PCR. As shown in Table 2, the highest expression was seen in the human osteosarcoma line MG63 where expression relative to HPRT ranged from 14 to 65 in different experiments. MG63 cells showed more than 1000-fold greater expression than HEK293 cells (expression relative to HPRT between 0.01 and 0.06). The human breast cancer cell lines MDA-MB-231 and MCF7 also expressed much lower *FBN1* mRNA levels than MG63 cells (expression relative to HPRT approximately 0.20; Table 2).

Since there was detectable *FBN1* mRNA expression in all cell types tested, we also assessed the production of fibrillin-1 protein in these cell lines using immunocytochemistry with a mouse anti-fibrillin-1 monoclonal antibody [22]. As shown in Fig. 3A, at day 14 after plating the majority of the staining in MG63 cultures was present in the extracellular region, in fibres similar to those seen in skin and aortic fibroblast cultures (for example, see [2,22–24]). In contrast, at 14 days after plating HEK293 cells showed minimal staining, with no fibres (Fig. 3B). This suggests that the mRNA expression seen in HEK293 cells (Table 2) does not result in production of fibrillin-1 protein that is incorporated into the extracellular matrix. The human breast cancer cell lines MCF7 and MDA-MB-231 (not shown), showed similar results to HEK293, consistent with quantitative reverse transcriptase PCR results (Table 2).

Assessment of human FBN1 promoter activity

Our bioinformatic study (based on CAGE analysis and sequence conservation) identified the TSS region which should be preceded by a promoter. The experimental study showed that the cell line MG63 expressed *FBN1* mRNA and produced a matrix containing fibrillin-1, while the three other cell lines (including the line HEK293 used in the previous study [12]) had low levels of mRNA and protein. We therefore wished to determine whether the putative promoter region had lineage-specific promoter activity. A 662 bp sequence including the full conserved TSS region of the human *FBN1* gene (as identified by CAGE analysis) was amplified by PCR and inserted into the firefly luciferase expression vector pGL2B, upstream of the luciferase coding sequence. This construct covered bases −1351 to −690 numbered from the putative start codon (based on GenBank entry L19896). It therefore included the experimentally-verified major TSS cluster. As shown in Fig. 4, this promoter construct (pUQ42) had detectable promoter activity in all of the human cell lines studied. Expressed relative to the pGL2B basal activity, the promoter activity of pUQ42 was 5–10 fold greater in MG63 cells than HEK293 cells or the human breast cancer lines (Fig. 4), These data are consistent with the higher level of mRNA expression in MG63 cells (Table 2). Deletion of 104 bases from the 5′ end (pUQ12) made little difference to the activity in most cell lines, but deletion of 236 bases from the 5′ end (pUQ09) reduced the activity considerably in MG63, HEK and MCF7 cells (Fig. 4). A construct with an internal deletion of one of the two putative E2F sites also showed reduced activity in most experiments for all cell lines (Fig. 4). When expressed relative to the activity of a positive control, the SV40-based pGL2C (Fig. 4 legend), *FBN1* promoter activity of pUQ42 was three times higher in MG63 than in the breast cancer lines but this comparison could not be made for HEK293 cells because activity of the positive control was very low.

Discussion

This study used publicly available genomic resources to assess the promoter region of *FBN1*, the gene which is associated with the human genetic disease Marfan syndrome. The analysis of TSS use for both mouse and human supported the view that the *FBN1* gene has a single dominant promoter, and hence that Exon A is the dominant 5′-non-coding exon of the gene. For humans, transcripts containing Exon B or C probably derive from the background low-level TSSs scattered over the 5′ region (Fig. 1), with transcripts containing weak splice donors that enable splicing into the strong splice acceptor upstream of Exon 1. Indeed, sequences flanking putative Exon B had minimal promoter activity in reporter assays [12]. There are also likely to be additional alternative 5′ exons. For example, a number of singleton TSSs were detected by CAGE analysis. Singletons usually represent real TSSs [25], so greater depth of sampling would probably validate these sites in *FBN1* and identify others. There is a similar complex pattern of

alternative exons in the trophoblast-specific promoter region of the *Csf1r* locus in mouse [26].

The results identified the *FBN1* promoter as a broad class promoter, with a number of closely located TSS rather than a distinct TSS at a single position [27]. As seen here, broad class promoters commonly correspond to CpG islands. Genome wide analysis [16] has shown that the dominant TSSs within a broad promoter are preceded by a GC-rich element that binds SP1, around 40–50 bp upstream. Our study identified this element in the *FBN1* promoter, highly conserved across mammalian species (Supplementary Fig. 2). Around 20 bp downstream of the major TSS of a broad promoter, genome analysis has identified another GC-rich element, with a characteristic GCG trinucleotide core [16], echoed by a third element about 10 bp further downstream. This structure was also found in the *FBN1* promoter region (Supplementary Fig. 2). The placement of these GCG elements might explain the preferential use of alternative TSS between the species.

In mouse, *Fbn1* is expressed at high levels in proliferating mesenchymal cells and tissues, regardless of their lineage commitment (smooth, cardiac and skeletal muscle, adipocytes, endothelial cells, osteoblasts, fibroblasts and chondrocytes) and absent from haemopoietic cells (see BioGPS). We therefore re-examined the promoter region for candidate conserved sequence elements that might explain this pattern of expression and detected a number of relevant transcription factor binding motifs (Table 1). We then scanned the literature to identify transcription factors that have been shown to bind these motifs and have also been implicated in mesenchyme lineage determination (references in Table 1). The survey of transcription factor binding sites in a previous report [12] did not take account of the pattern of expression of *FBN1*. For example, a binding motif for AP-2 was highlighted, but this is unlikely to be functional since the AP-2 family is associated with gene regulation in ectoderm and neural crest lineages [28] where *FBN1* is not expressed. The previous study highlighted a motif for the NFI family of factors which has been implicated in development of the nervous system. This motif may be relevant, since one member, Nfix, is largely co-expressed with *Fbn1* in mouse mRNA (BioGPS database and Summers et al., in preparation) and the mouse knockout had skeletal and muscle abnormalities as well as neurological effects [29]. Amongst the candidates we identified, the predicted GATA motifs were previously shown to be essential for promoter activity in HEK293 cells [12]. Further work is required to assess whether all the detected motifs are active.

The previous study of the *FBN1* promoter region [12] did not identify the TSSs experimentally. We found no TSSs detected by CAGE analysis at the Inr they proposed from sequence analysis [12] (see Fig. 2) and TSS distribution suggested that the major Inr is downstream of that site. Our analysis of TSSs for *FBN1*, also revealed that the active promoter constructs in the previous study [12] lacked key elements of the promoter region. As shown in the diagram in Fig. 4, construct P1, the most active construct of that study, did not contain the major TSS region identified by CAGE. Construct P2, with approximately half the activity of P1, contained the TSS region but lacked most of the 5′ sequence which is likely to be necessary for full promoter activity. The most active construct used in our study contained both the TSS region and the 5′ sequence of P1 (Fig. 4). In addition, as demonstrated by our results, the previous study [12] examined promoter activity in a cell line, HEK293, that makes little *FBN1* mRNA (Table 2) and does not make a fibrillin-1 matrix (Fig. 3). The level of *FBN1* mRNA in MG63 cells was at least 1000 times greater than in HEK293 cells (Table 2). This difference is consistent with expression in corresponding primary cells. MG63 is an osteosarcoma and *FBN1* mRNA level is high in primary osteoblasts and bone [14]. Conversely, HEK293 is probably neuronal [30], and the *FBN1* gene is not expressed in neuronal tissues or primary cells (data from BioGPS and RefExA).

Despite the massive differential expression of *FBN1* mRNA, promoter activity for the highly conserved proximal promoter region in MG63 cells was only ~5-fold higher than in the two breast cancer lines, relative to either the pGL2B empty vector, or the pGL2C SV40 positive control (Fig. 4). In HEK293 cells, the SV40 promoter was less active than the *FBN1* promoter. HEK cells were created by transformation with adenovirus, and the adenoviral E1A gene product represses transcription from a number of promoters including SV40 [28], accounting for the low level of luciferase activity in HEK cells transfected with pGL2C. The E1A product expressed in HEK293 cells could act in part through the E2F motif of the *FBN1* promoter, a well-documented target of E1A [29]. Given the presence of an E2F site, the dependence of activity on that site (Fig. 4 and [12]), and the fact that extracellular matrix production is itself associated with cell proliferation, the activity of the promoter in *FBN1*-expressing as well as non-expressing cells may be controlled by the cell cycle. Distal elements upstream and downstream of the conserved promoter may contribute to tissue specificity. The ECR Browser revealed additional regions of substantial homology upstream of the Exon A promoter. MG63 cells will be an appropriate cellular system in which to determine the role of these elements in expression of *FBN1*.

Variations in promoter activity can have a heritable effect on gene expression [9,10]. Thus polymorphic variation in the critical elements of the *FBN1* promoter may be involved in controlling the absolute level of *FBN1* mRNA and hence protein. Once these elements and the transcription factors which bind to them have been determined, studies in patient and control cells will be required to assess the significance in determining the severity and variability of the phenotype in Marfan syndrome.

Materials and methods

Bioinformatic analysis of the FBN1 promoter region in mammals

Microarray data in the public domain (BioGPS, RefExA, GeoProfiles) were searched to establish cell types with high and low expression levels of *FBN1* mRNA. The human and mouse *FBN1* transcriptional start sites (TSS) were identified from the FANTOM3 CAGE database [15,25,30]. *FBN1* homologues in 28 mammalian genomes were retrieved from the Ensembl trace sequence database, using the program SSAHA to identify matches with the 100 bp region containing the TSS and 100 bases flanking it on both sides. *Macropus eugenii* sequences were from The Wallaby Genome Project database. Multiple alignments were performed using ClustalW2 [31] and corrected manually using the alignment editor associated with the program.

An analysis of evolutionarily conserved sites in the promoter region was made with the ECR Browser (comparing mouse, human, rhesus macaque and dog genomes). The program rVISTA 2.0 was then used to identify transcription factor binding motifs within the conserved regions (based on TRANSFAC Professional V10.2 library [32]) and these results were correlated with those derived using the JASPAR database [33]. Nucleosome phasing prediction for the conserved region was made using the RECON program [21].

Cells and cell culture

Human MG63 osteosarcoma cells and human MCF7 and MDA-MB-231 breast cancer cells were grown in minimal essential medium (MEM) with 10% heat inactivated fetal calf serum, 2 mM L-glutamine (Glutamax-1), 100 μg/mL streptomycin sulphate and 100 U/mL penicillin. MG63 and MCF7 cells were supplemented with 0.1 mM non-essential amino acids and 1.0 mM sodium pyruvate. MCF7 cells were also supplemented with 10 μg/mL bovine insulin (Sigma-Aldrich, Castle Hill, NSW, Australia). Human embryonic kidney cells (HEK-293) were grown in Dulbecco's modified Eagle's medium

(DMEM) in the presence of fetal calf serum, L-glutamine, penicillin and streptomycin as above. Media components were from Invitrogen Australia, Mt Waverly, Victoria, Australia, except where specified. The cells were maintained in a humidified incubator at 37 °C with 5% CO_2.

FBN1 mRNA and protein expression analysis

RNA was harvested from cells that were confluent at the time of harvest, 7 days after seeding into T75 flasks, using the RNeasy Mini Kit (Qiagen Inc, Hilden, Germany). cDNA was synthesised in duplicate using oligo-dT primers (Geneworks, Adelaide, South Australia, Australia) and Superscript reverse transcriptase (Invitrogen Australia). Quantitative real time PCR was performed in triplicate for each cDNA using SYBRGreen PCR mastermix (Applied Biosystems, Scoresby, Victoria, Australia) in an ABI Prism 7000 Sequence Detection System (Applied Biosystems) according to the manufacturer's instructions. Expression was calculated relative to hypoxanthine guanine phosphoribosyl transferase. Primer sequences are available in Supplementary Table 1. Fibrillin-1 protein was detected using a monoclonal antibody by immunocytochemistry as described previously [22].

FBN1 promoter analysis

Primers were designed using the program Primer3 to amplify the region identified to contain the major TSS for human *FBN1*. Eight nucleotides containing the recognition site for the restriction enzyme *Bgl*II were added to the primers at the 5′ end. Primer sequences for all constructs and their location relative to the putative ATG start codon are available in Supplementary Table 1. PCR fragments were cloned into the T-tailed vector pGEM T-easy (Promega Corp, Madison, WI, USA). Fragments were excised from the vector using *Bgl*II and ligated into the phosphatase treated *Bgl*II site of the luciferase reporter vector pGL2B (Promega). Deletion of 8 bases within the putative promoter region was accomplished using a site directed mutagenesis protocol [34].

Plasmids were prepared using an endotoxin-free maxiprep kit (Qiagen). Twenty-four hours prior to transfection, a total of 10^5 cells were plated into wells of a 12 well plate. Antibiotics were removed from the medium three hours prior to transfection. Prior to transfection, the medium was replaced with Opti-MEM (antibiotic and serum free). 400 ng of the pGL2B construct was transfected using LipofectAMINE 2000™ (Invitrogen Australia) following the manufacturer's instructions. Four hours after adding the transfection mix, medium was replaced with DMEM containing 10% fetal calf serum and appropriate supplements. Twenty-four hours after transfection, the cells were washed with phosphate buffered saline, lysed with 120 μL lysis buffer (0.5 M HEPES at pH 7.9, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 0.1% triton N-101) and frozen at −80 °C. The activity of luciferase was measured in 50 μL of lysate using the Luciferase Reporter Gene Assay High Sensitivity (Roche Applied Science, Castle Hill, NSW, Australia) in a 1450 Microbeta TriLux luminometer (PerkinElmer, Fremont, CA, USA). Luciferase activity was corrected for protein concentration assayed using the Bradford reagent (BioRad Laboratories, Gladesville, NSW, Australia). All experiments were performed in triplicate.

Web sources

BioGPS The Gene Portal Hub, http://biogps.gnf.org
BioGPS *FBN1* entry, http://biogps.gnf.org/#goto=genereport&id=2200
ClustalW2, http://www.ebi.ac.uk/Tools/clustalw2/index.html
ECR Browser, http://ecrbrowser.dcode.org/
Ensembl Database, http://www.ensembl.org
Ensembl Trace Sequence Database, http://trace.ensembl.org/
FANTOM3 CAGE Database, http://fantom.gsc.riken.jp/3/
GenBank, http://www.ncbi.nlm.nih.gov/sites/entrez?db=nucleotide
GeoProfiles, Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/
JASPAR Transcription Factor Motif Database, http://jaspar.cgb.ki.se/
Primer3 primer design program, http://frodo.wi.mit.edu/
RefExA, Laboratory for Systems Biology and Medicine, http://www.lsbm.org
RECON, http://wwwmgs.bionet.nsc.ru/mgs/programs/recon/
rVISTA, http://rvista.dcode.org/
Wallaby Genome Project, http://www.hgsc.bcm.tmc.edu/project-species-m-Wallaby.hgsc?pageLocation=Wallaby

Acknowledgments

We thank Fiona Dowell and Truc Linh Tran (School of Chemistry and Molecular Biosciences, The University of Queensland) for preliminary experiments and Carl Morrow (School of Chemistry and Molecular Biosciences, The University of Queensland) for technical assistance. This work was supported by a grant-in-aid from the National Heart Foundation of Australia (G06B2527) and by the Biotechnology and Biological Sciences Council of the United Kingdom (Institutional Strategic Programme Grant to The Roslin Institute).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2009.06.005.

References

[1] H.C. Dietz, et al., Marfan phenotype variability in a family segregating a missense mutation in the epidermal growth factor-like motif of the fibrillin gene, J. Clin. Invest. 89 (1992) 1674–1680.
[2] K.M. Summers, et al., An integrated approach to management of Marfan syndrome caused by an FBN1 exon 18 mutation in an Australian Aboriginal family, Clin. Genet. 65 (2004) 66–69.
[3] R.E. Pyeritz, The Marfan syndrome, Annu. Rev. Med. 51 (2000) 481–510.
[4] K.M. Summers, J.A. West, M.M. Peterson, D. Stark, J.J. McGill, M.J. West, Challenges in the diagnosis of Marfan syndrome, Med. J. Aust. 184 (2006) 627–631.
[5] S. Hutchinson, et al., Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype, Hum. Mol. Genet. 12 (2003) 2269–2276.
[6] D.P. Judge, et al., Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome, J. Clin. Invest. 114 (2004) 172–181.
[7] G. Matyas, et al., Large genomic fibrillin-1 (FBN1) gene deletions provide evidence for true haploinsufficiency in Marfan syndrome, Hum. Genet. 122 (2007) 23–32.
[8] K.K. Singh, P.C. Shukla, K. Rommel, J. Schmidtke, M. Arslan-Kirchner, Sequence variations in the 5′ upstream regions of the FBN1 gene associated with Marfan syndrome, Eur. J. Hum. Genet. 14 (2006) 876–879.
[9] D.J. de Koning, C.S. Haley, Genetical genomics in humans and model organisms, Trends Genet. 21 (2005) 377–381.
[10] N. Hubner, et al., Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease, Nat. Genet. 37 (2005) 243–253.
[11] G.M. Corson, S.C. Chalberg, H.C. Dietz, N.L. Charbonneau, L.Y. Sakai, Fibrillin binds calcium and is coded by cDNAs that reveal a multidomain structure and alternatively spliced exons at the 5′ end, Genomics 17 (1993) 476–484.
[12] G. Guo, S. Bauer, J. Hecht, M.H. Schulz, A. Busche, P.N. Robinson, A short ultraconserved sequence drives transcription from an alternate FBN1 promoter, Int. J. Biochem. Cell Biol. 40 (2008) 638–650.
[13] K.K. Singh, P.C. Shukla, J. Schmidtke, Conservation of 5′-upstream region of the FBN1 gene in primates, Eur. J. Hum. Genet. 16 (2008) 869–872.
[14] S. Kitahama, et al., Expression of fibrillins and other microfibril-associated proteins in human bone and osteoblast-like cells, Bone 27 (2000) 61–67.
[15] P. Carninci, et al., The transcriptional landscape of the mammalian genome, Science 309 (2005) 1559–1563.
[16] M.C. Frith, E. Valen, A. Krogh, Y. Hayashizaki, P. Carninci, A. Sandelin, A code for transcription initiation in mammalian genomes, Genome Res. 18 (2008) 1–12.
[17] M. Gardiner-Garden, M. Frommer, CpG islands in vertebrate genomes, J. Mol. Biol. 196 (1987) 261–282.
[18] M. Bieda, X. Xu, M.A. Singer, R. Green, P.J. Farnham, Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome, Genome Res. 16 (2006) 595–605.
[19] R. Vernell, K. Helin, H. Muller, Identification of target genes of the p16INK4A-pRB-E2F pathway, J. Biol. Chem. 278 (2003) 46124–46137.
[20] K. Blyth, E.R. Cameron, J.C. Neil, The RUNX genes: gain or loss of function in cancer, Nat. Rev. Cancer 5 (2005) 376–387.

[21] V.G. Levitsky, RECON: a program for prediction of nucleosome formation potential, Nucleic Acids Res. 32 (2004) W346–W349.
[22] M. Nataatmadja, et al., Abnormal extracellular matrix protein transport associated with increased apoptosis of vascular smooth muscle cells in Marfan syndrome and bicuspid aortic valve thoracic aortic aneurysm, Circulation 108 (Suppl 1) (2003) II329–II334.
[23] G.B. Schaefer, M. Godfrey, Quantitation of fibrillin immunofluorescence in fibroblast cultures in the Marfan syndrome, Clin. Genet. 47 (1995) 144–149.
[24] K.M. Summers, et al., Histopathology and fibrillin-1 distribution in severe early onset Marfan syndrome, Am. J. Med. Genet. A 139 (2005) 2–8.
[25] P. Carninci, et al., Genome-wide analysis of mammalian promoter architecture and evolution, Nat. Genet. 38 (2006) 626–635.
[26] R.T. Sasmono, et al., A macrophage colony-stimulating factor receptor-green fluorescent protein transgene is expressed throughout the mononuclear phagocyte system of the mouse, Blood 101 (2003) 1155–1163.
[27] A. Sandelin, P. Carninci, B. Lenhard, J. Ponjavic, Y. Hayashizaki, D.A. Hume, Mammalian RNA polymerase II core promoters: insights from genome-wide studies, Nat. Rev. Genet. 8 (2007) 424–436.
[28] C. Rochette-Egly, C. Fromental, P. Chambon, General repression of enhanson activity by the adenovirus-2 E1A proteins, Genes Dev. 4 (1990) 137–150.
[29] A.J. Berk, Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus, Oncogene 24 (2005) 7673–7685.
[30] V.B. Bajic, et al., Mice and men: their promoter properties, PLoS Genet. 2 (2006) e54.
[31] M.A. Larkin, et al., Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.
[32] V. Matys, et al., TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes, Nucleic Acids Res. 34 (2006) D108–D110.
[33] D. Vlieghe, et al., A new generation of JASPAR, the open-access repository for transcription factor binding site profiles, Nucleic Acids Res. 34 (2006) D95–D97.
[34] K.J. Chappell, M.J. Stoermer, D.P. Fairlie, P.R. Young, Insights to substrate binding and processing by West Nile Virus NS3 protease through combined modeling, protease mutagenesis, and kinetic studies, J. Biol. Chem. 281 (2006) 38448–38458.
[35] C. Nerlov, The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control, Trends Cell Biol. 17 (2007) 318–324.
[36] C.S. Shin, et al., CCAAT/enhancer-binding protein delta activates the Runx2-mediated transcription of mouse osteocalcin II promoter, J. Mol. Endocrinol. 36 (2006) 531–546.
[37] S.J. Greco, K. Liu, P. Rameshwar, Functional similarities among genes regulated by OCT4 in human mesenchymal and embryonic stem cells, Stem Cells 25 (2007) 3143–3154.
[38] M. Nishimoto, S. Miyagi, T. Katayanagi, M. Tomioka, M. Muramatsu, A. Okuda, The embryonic Octamer factor 3/4 displays distinct DNA binding specificity from those of other Octamer factors, Biochem. Biophys. Res. Commun. 302 (2003) 581–586.
[39] M. Buckingham, F. Relaix, The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions, Annu. Rev. Cell Dev. Biol. 23 (2007) 645–673.
[40] F.J. Naya, E. Olson, MEF2: a transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation, Curr. Opin. Cell Biol. 11 (1999) 683–688.
[41] D. ten Berge, A. Brouwer, J. Korving, J.F. Martin, F. Meijlink, Prx1 and Prx2 in skeletogenesis: roles in the craniofacial region, inner ear and limbs, Development 125 (1998) 3831–3842.
[42] P.Y. Jay, et al., Impaired mesenchymal cell function in Gata4 mutant mice leads to diaphragmatic hernias and primary lung defects, Dev. Biol. 301 (2007) 602–614.
[43] C. Mauritz, et al., Generation of functional murine cardiac myocytes from induced pluripotent stem cells, Circulation 118 (2008) 507–517.
[44] S. Polager, D. Ginsberg, E2F — at the crossroads of life and death, Trends Cell Biol. 18 (2008) 528–535.
[45] B. Scheijen, M. Bronk, T. van der Meer, R. Bernards, Constitutive E2F1 overexpression delays endochondral bone formation by inhibiting chondrocyte differentiation, Mol. Cell. Biol. 23 (2003) 3656–3668.
[46] M.M. Cohen Jr., The new bone biology: pathologic, molecular, and clinical correlates, Am. J. Med. Genet. A 140 (2006) 2646–2706.
[47] K. Seki, et al., Mouse Snail family transcription repressors regulate chondrocyte, extracellular matrix, type II collagen, and aggrecan, J. Biol. Chem. 278 (2003) 41862–41870.
[48] H.H. Arnold, B. Winter, Muscle differentiation: more complexity to the network of myogenic regulators, Curr. Opin. Genet. Dev. 8 (1998) 539–544.